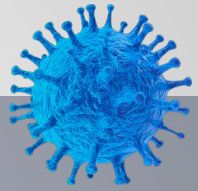
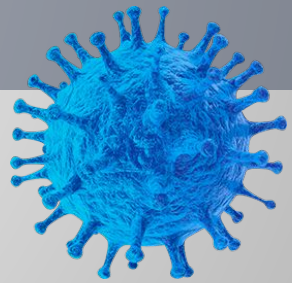
Home Value Patterns Due to COVID-19

Data Analytics by: Jason Sheridan, Anna Sours, Jenni Davis, Austin Olea, & Mark Blankenship






Hypothesis

Where COVID-19 cases are higher in counties with major cities, home values will increase in the surrounding counties.

Questions:

- How are people migrating throughout the US since COVID 19 began?
- Has this impacted home values?
- Is there a relationship between cases and the housing market?

Data Sources:

COVID-19 Cases:

- NY Times US states & counties
  - Texts files with state, county, date, cases, etc.
- US census data
  - Shapefiles
- API calls
  - Coordinates
- New dependencies for mapping
  - geopandas
  - folium

Home values:

- Zillow Home Value Index (ZHVI) from zillow.com
  - CSV files updated monthly
  - ZHVI values for Metro & U.S.
  - Monthly values from 1/31/1996 - 12/31/2020

COVID-19 Data Exploration

Analysis 1:

- NY Times US states & counties
  - **Q: What states have the highest count of COVID-19 cases? And of those states, which counties had the highest count of COVID-19 cases?**

Exploration & Data Wrangling:

- NY Times data
  - US states & counties data - date, state, county, cases, etc.

Methods:

- ➔ **Filtering**
- ➔ **Grouping**
- ➔ **Stats**
- ➔ **Dropping values**
- ➔ **Sorting values**
- ➔ **Horizontal bar chart**

Analysis 1: NYTimes Data

```
# Store filepath in a variable
file_one = "output_data/us-states.txt"
us_states = pd.read_csv(file_one)
file_two = "output_data/us-counties.txt"
us_counties = pd.read_csv(file_two)
```

```
#Use groupby method to isolate states
grouped = us_states.groupby(['state'])
```

```
#Calculate maximum number of cases
max_cases = grouped['cases'].max()
```

```
#Create a data frame
remove_terr = pd.DataFrame(max_cases)

#Remove US territories
fifty_states = remove_terr.drop(['Guam', 'Puerto Rico', 'Northern Mariana
```

```
#Sort cases by highest to lowest number by state
fifty_states = fifty_states.sort_values(['cases'],ascending='False')
```

```
#Create a bar chart to find top three states with highest amount of cases
panda_chart = fifty_states.plot(kind='barh', figsize=(15, 10))
panda_chart.set_xlabel("Total Number")
panda_chart.set_ylabel("States")
plt.title("States and Maximum Cases")
# plt.tight_layout()
plt.show()
```

COVID-19 Data Exploration

Analysis 2:

- API calls, US census data & geopandas
  - **Q: Can we take the NYTimes data and layer it onto a map of the United States and have it display state name and maximum case counts?**

Exploration & Data Wrangling:

- API calls
  - Coordinates
- US census data
  - Shapefiles
- Geopandas

Methods:

- ➔ **API calls** 😑
- ➔ **Shapefiles**
- ➔ **Export/Import csv files**
- ➔ **More: filtering, grouping, dropping**
- ➔ **Adding/renaming columns**
- ➔ **Merging**

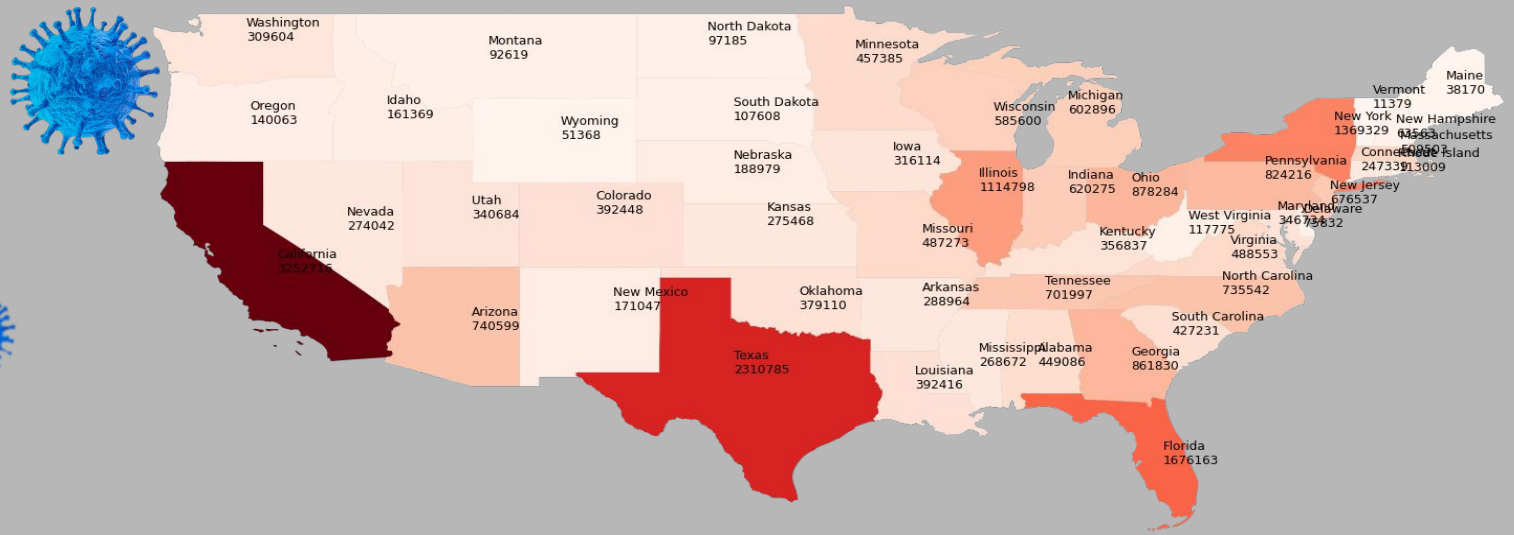
```
In [2]: # List for holding lat_lngs and list of cities to be found
        lat_lngs = []
        cities = ['Los Angeles', 'Riverside', 'Victorville','Bakersfield','Cartago','Thousand Oaks','Irvine','Escondido',
                  'Brawley','Houston','Conroe','Dayton','Beach City','League City','Angleton','Rosenberg','Hempstead',
                  'Dallas','Fort Worth','Weatherford','Decatur','Denton','McKinney','Cleburne','Palmer','Terrell','Rockwall',
                  'Miami','Davie','West Palm Beach','Key West','Golden Gate','Fort Myers','Clewiston','Buckhead Ridge',
                  'Taylor Creek','Palm City']

        # Create a set of random lat and lng combinations
        lats = np.random.uniform(lat_range[0], lat_range[1], size=40)
        lngs = np.random.uniform(lng_range[0], lng_range[1], size=40)
        lat_lngs = zip(lats, lngs)

        # Identify city for each lat, lng combination
        for lat_lng in lat_lngs:
            city = citipy.nearest_city(lat_lng[0], lat_lng[1]).city_name

        # Print the city count to confirm sufficient count
        len(cities)
```

```
1 #Read into shapefile from US census data
2 file_path = "mapping/us_nation/cb_2018_us_state_20m.shp"
3 usa_map = gpd.read_file(file_path)
4 # type(usa_map)
```

Analysis 2: API calls for coordinates

Analysis 3: Shapefiles & geopandas

Maximum COVID-19 Cases per State (showing lower forty-eight)



States determined for further analysis:

- California
- Texas
- Florida

COVID-19 Data Exploration

Analysis 3:

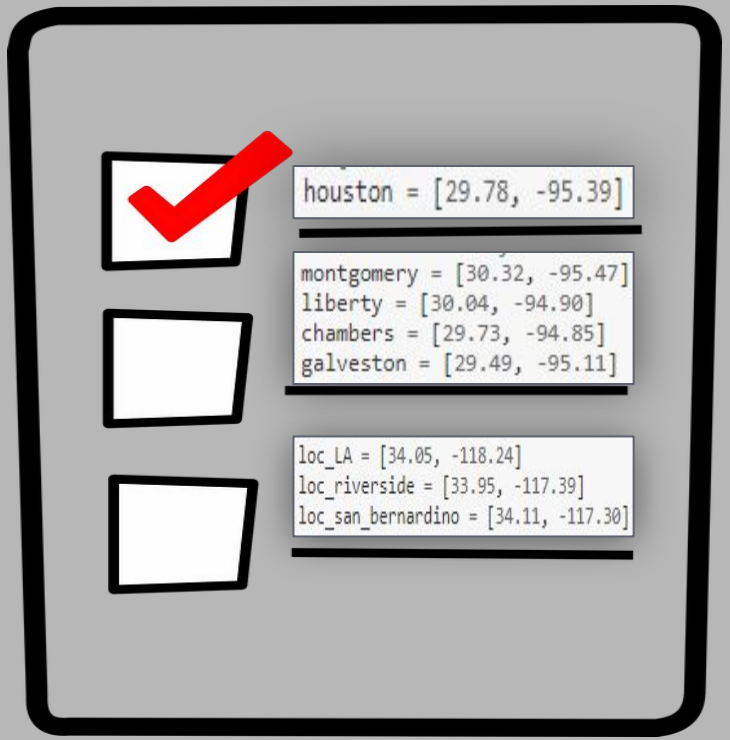
- More mapping...?
  - **Q: Can we use heat maps and markers to better highlight the counties with maximum cases and their surrounding counties?**

Exploration & Data Wrangling:

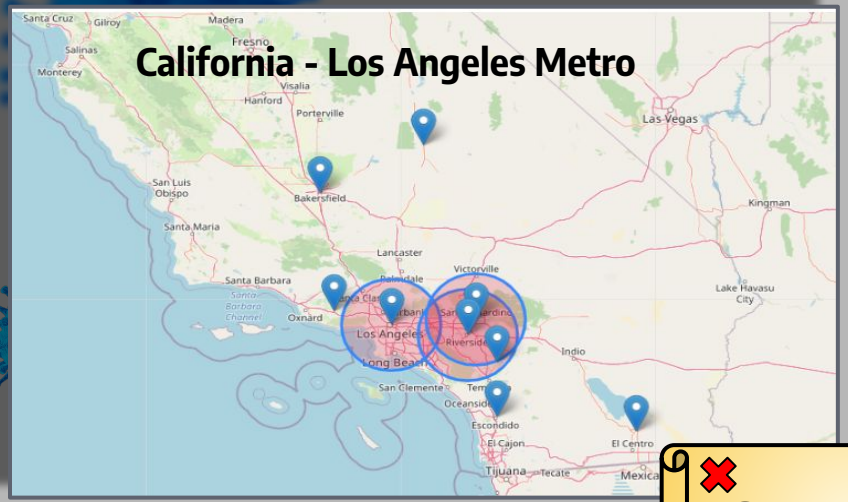
- Folium
  - Mapping library

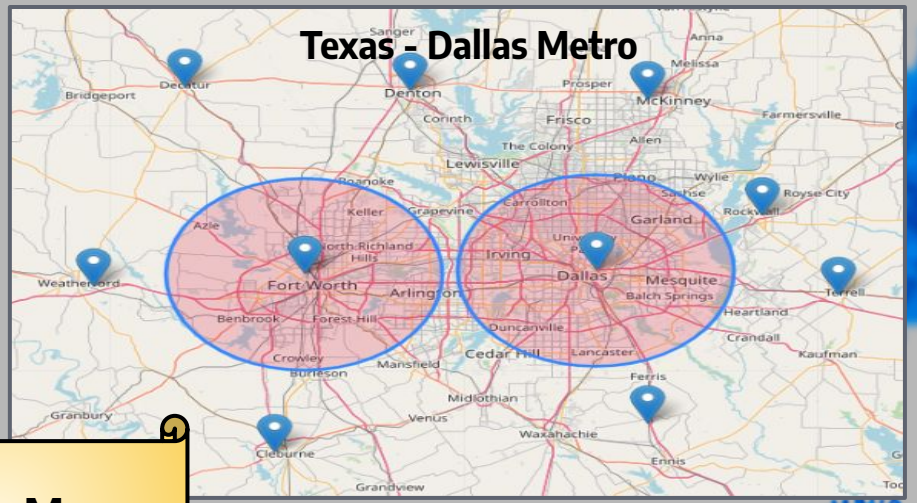
Methods:

- ~~Gmaps~~
- ~~Heatmaps~~
- **Folium**
- **Lists...**
- **Lists of coordinates...**
- **Lists of cities...**
- **LISTS!**

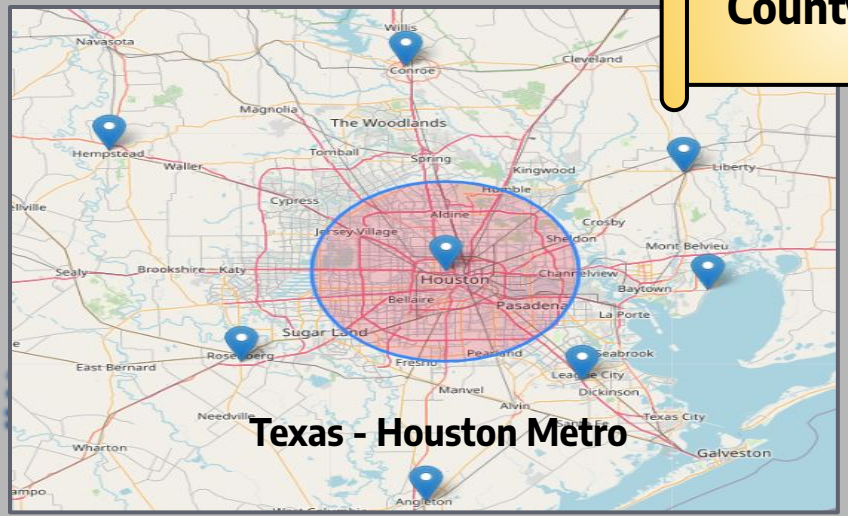


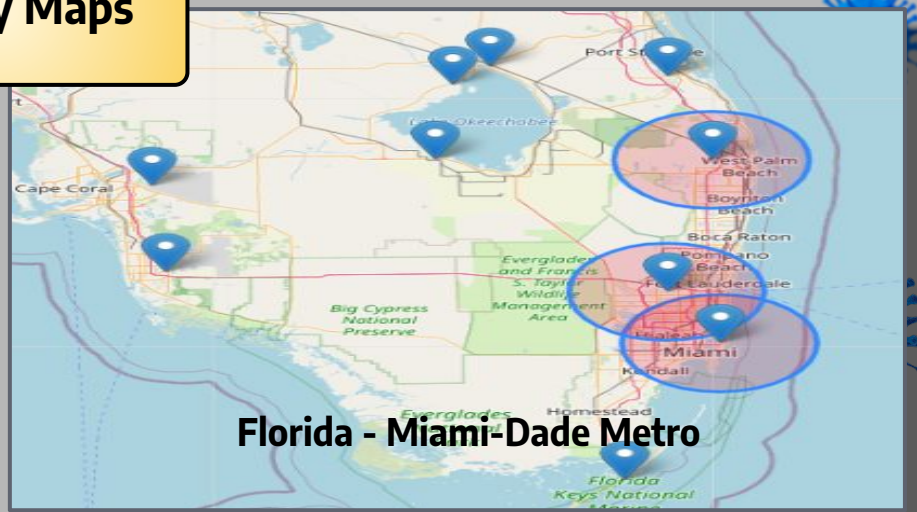
Analysis 4: County maps


California - Los Angeles Metro


Texas - Dallas Metro

County Maps


Texas - Houston Metro


Florida - Miami-Dade Metro

Home Value Data:

- **Q: What are home values in areas with highest count of COVID-19 cases?**
    - ZHVI reflects typical value and market changes for homes in the 35th to 65th percentile range in a given region
    - Narrowed down data to relevant years: 2017-2020
    - Dropped null values from numerical columns
    - No duplicate county names
    - Exported to CSV for use with analysis

In [158]:
```
# Read raw data

file_path = "zillow_data/zhvi_data/County_zhvi_uc_sfrcondo_tier_0.33_0.67_sm_sa_mon.csv"
home_values_df=pd.read_csv(file_path)

home_values_df.head()
```

Out[158]:

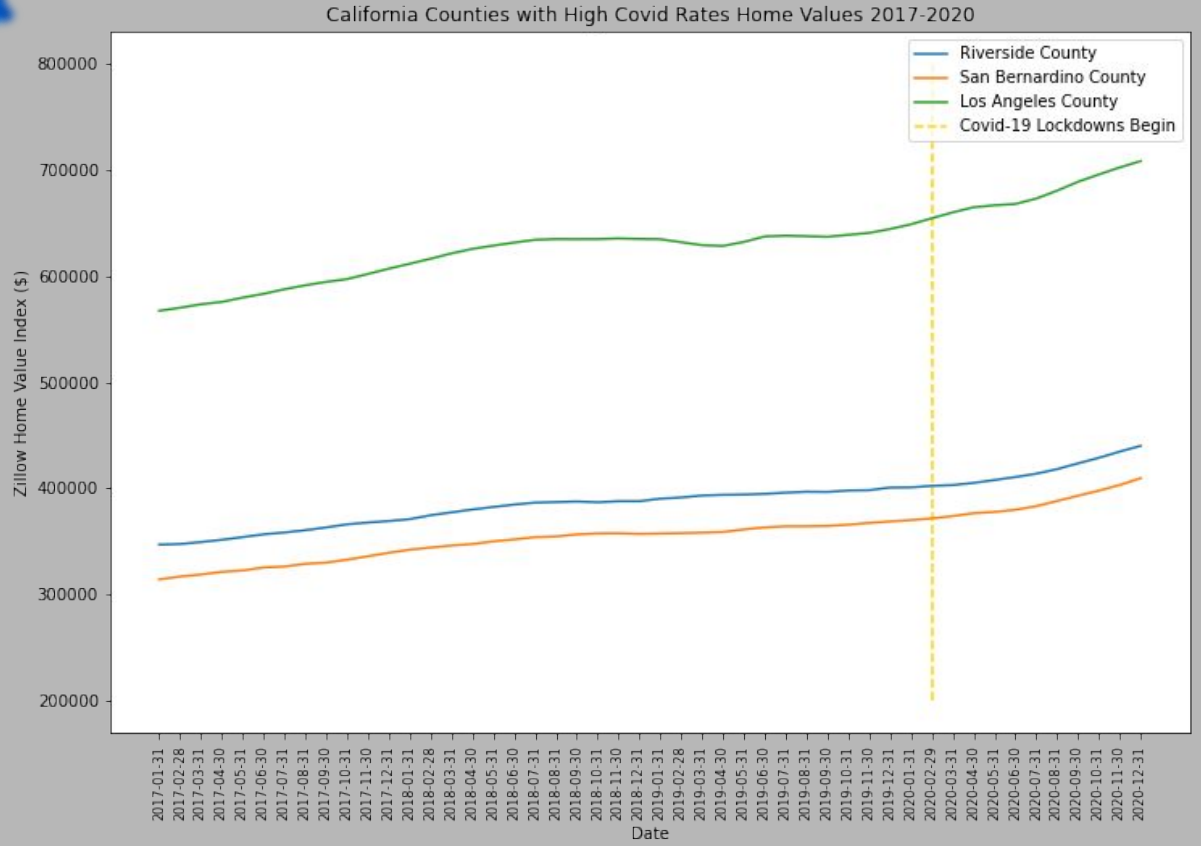
	RegionID	SizeRank	RegionName	RegionType	StateName	State	Metro	StateCodeFIPS	MunicipalCodeFIPS	1996-01-31	...	2020-03-31	2020-04-30	2020-0
0	3101	0	Los Angeles County	County	CA	CA	Los Angeles-Long Beach-Anaheim	6	37	177774.0	...	659773.0	664580.0	666444
1	139	1	Cook County	County	IL	IL	Chicago-Naperville-Elgin	17	31	154331.0	...	253684.0	254076.0	254595
2	1090	2	Harris County	County	TX	TX	Houston-The Woodlands-Sugar Land	48	201	99675.0	...	199807.0	200940.0	201912
3	2402	3	Maricopa County	County	AZ	AZ	Phoenix-Mesa-Scottsdale	4	13	116247.0	...	299209.0	302766.0	305996
4	2841	4	San Diego County	County	CA	CA	San Diego-Carlsbad	6	73	180788.0	...	615413.0	618539.0	622620

5 rows × 309 columns

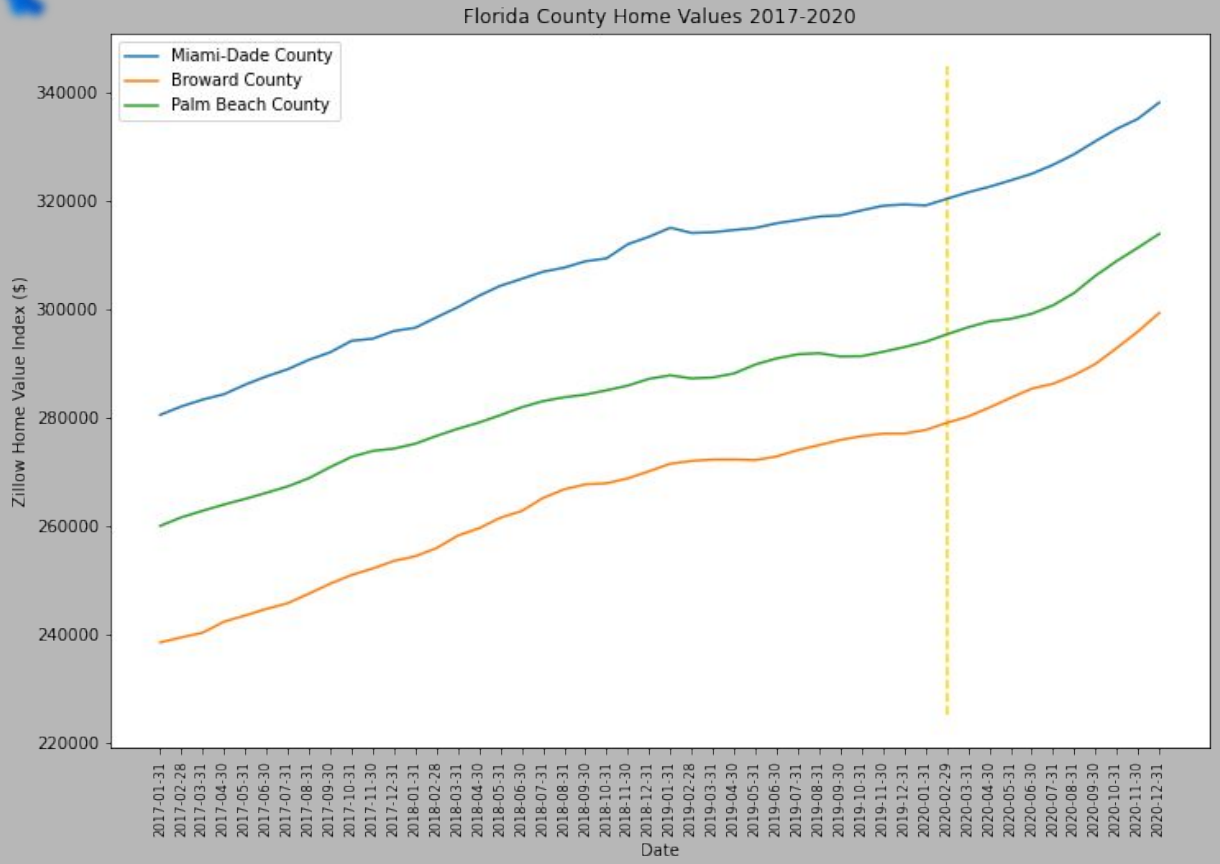
	RegionName	RegionType	State	Metro	StateCodeFIPS	MunicipalCodeFIPS	2017-01-31	2017-02-28	2017-03-31	2017-04-30	...	2020-03-31	2020-04-30	2020-05-31
0	Los Angeles County	County	CA	Los Angeles-Long Beach-Anaheim	6	37	567167.0	570020.0	573344.0	575480.0	...	659773.0	664580.0	666444.0
1	Cook County	County	IL	Chicago-Naperville-Elgin	17	31	235165.0	235814.0	237493.0	238830.0	...	253684.0	254076.0	254595.0
2	Harris County	County	TX	Houston-The Woodlands-Sugar Land	48	201	177989.0	178366.0	178869.0	179782.0	...	199807.0	200940.0	201912.0
3	Maricopa County	County	AZ	Phoenix-Mesa-Scottsdale	4	13	242028.0	243189.0	244568.0	245814.0	...	299209.0	302766.0	305996.0
4	San Diego County	County	CA	San Diego-Carlsbad	6	73	528758.0	531624.0	536251.0	542424.0	...	615413.0	618539.0	622620.0

5 rows × 54 columns

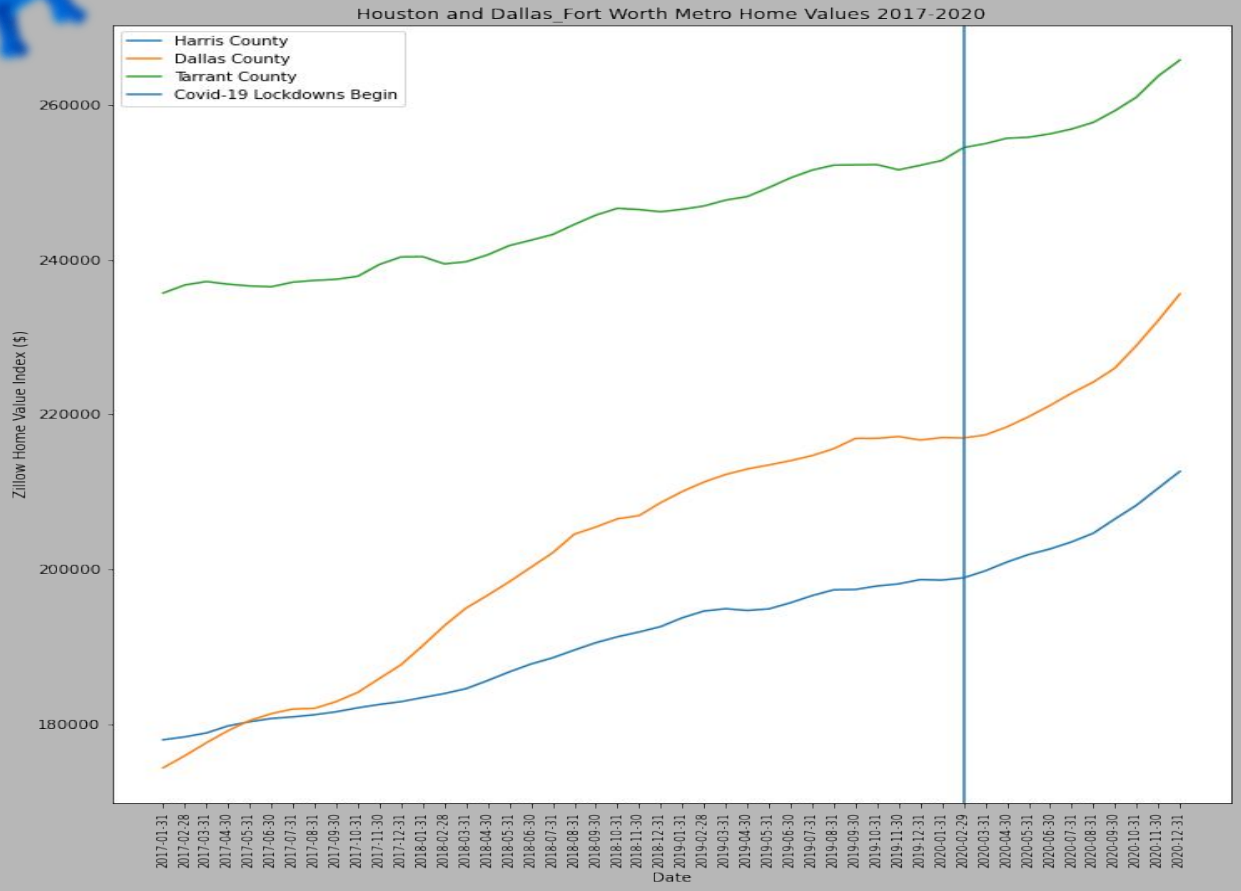
Home Value by County: California



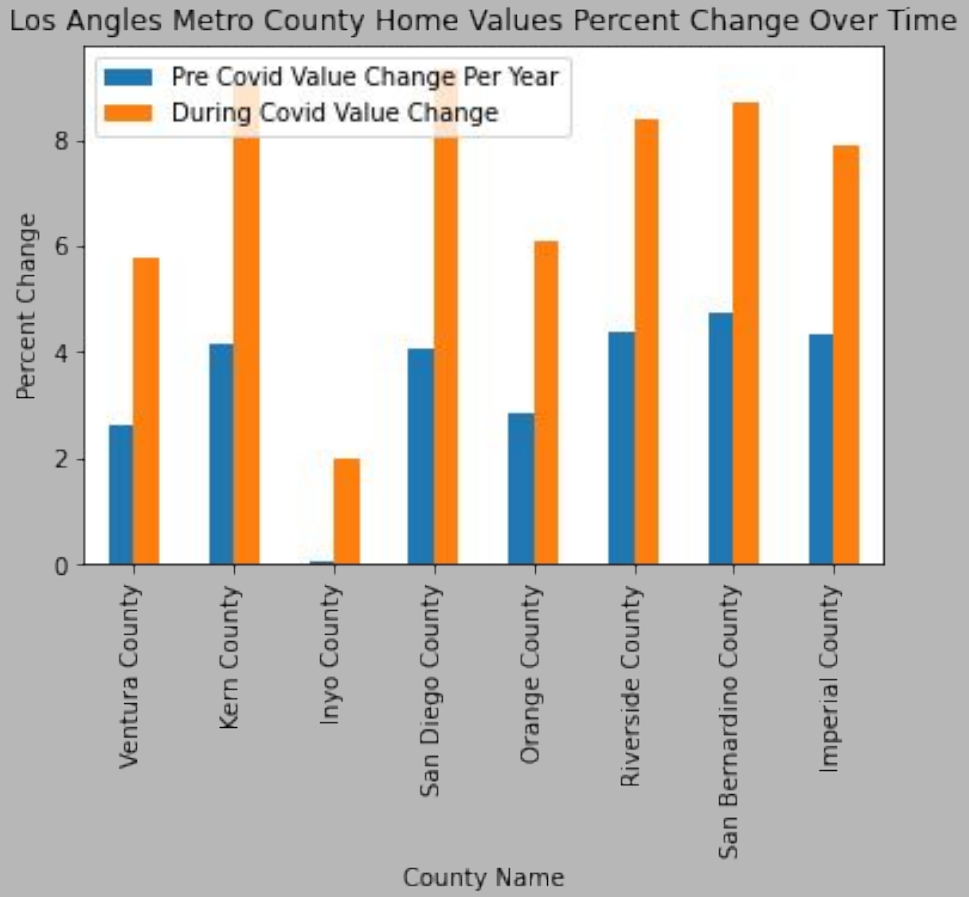
Home Value by County: Florida



Home Value by County: Texas

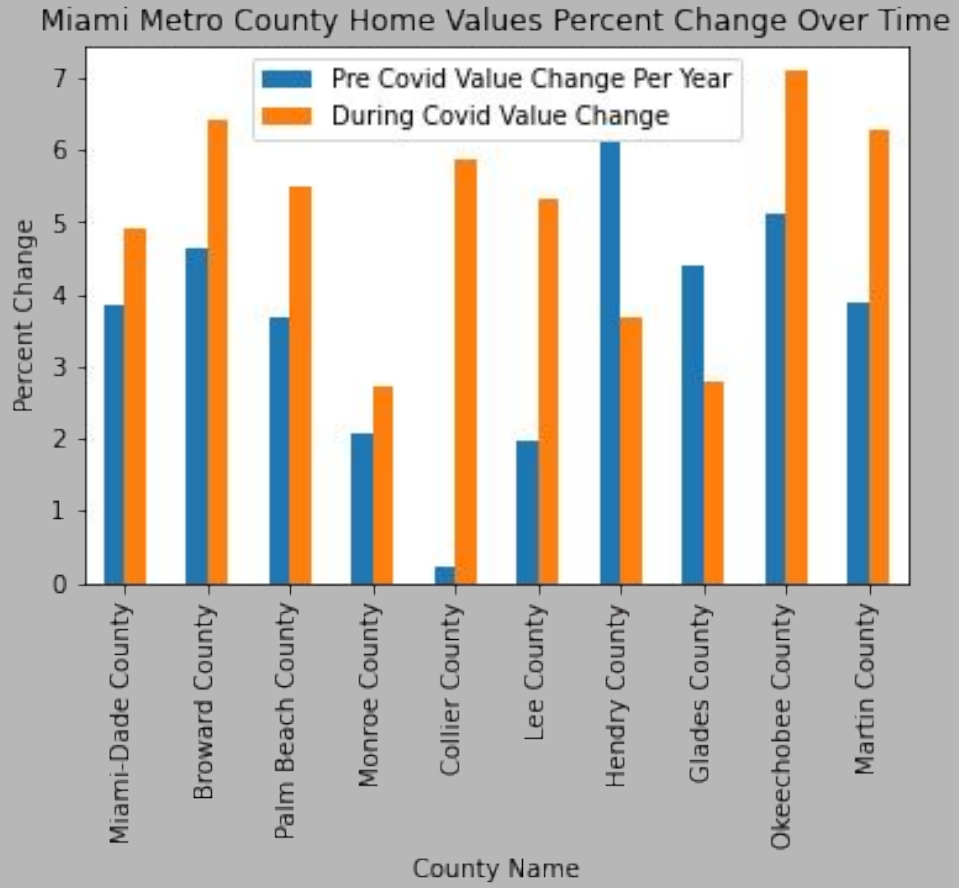


Home Value by County: California

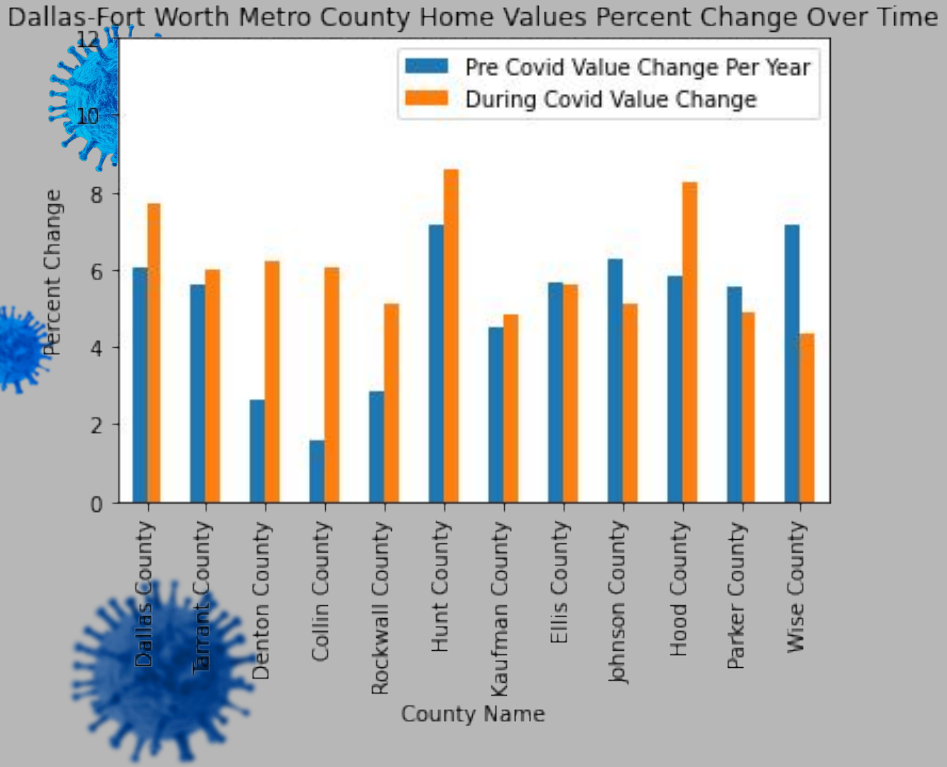


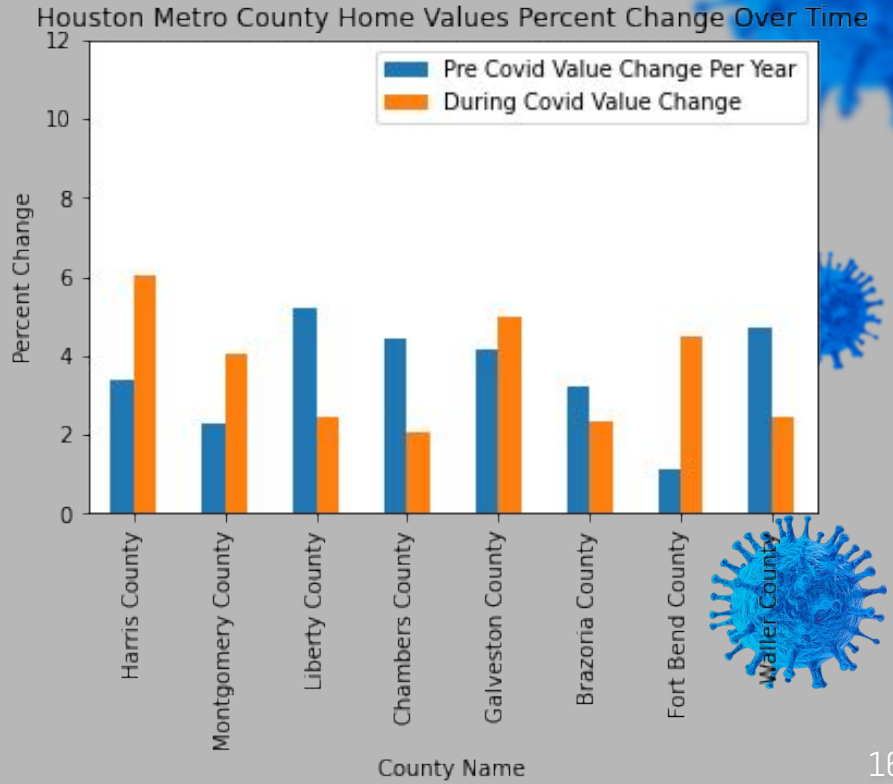
```
tx_percent_cols=tx_df[['RegionName','2017-03-31', '2020-02-29', '2020-03-31', '2020-12-31']]
Dallas_counties=["Dallas County", "Tarrant County", "Denton County", "Collin County", "Rockwall County", "Hunt County", "Kauf

pre_covid_dfw=[]
covid_dfw=[]


for each_county in Dallas_counties:

    tx_cnt_filter=tx_percent_cols['RegionName']==each_county  #just change county name and run cell

    mar2017=tx_percent_cols.loc[tx_cnt_filter,'2017-03-31']
    feb2020=tx_percent_cols.loc[tx_cnt_filter,'2020-02-29']
    mar2020=tx_percent_cols.loc[tx_cnt_filter,'2020-03-31']
    dec2020=tx_percent_cols.loc[tx_cnt_filter,'2020-12-31']


    pre_covid_annual=((((feb2020-mar2017)/feb2020)/3)*100).values[0]
    covid_annual=(((dec2020-mar2020)/dec2020)*100).values[0]

    pre_covid_dfw.append(pre_covid_annual)
    covid_dfw.append(covid_annual)
```

Home Value by County: Florida



Home Value By County: Texas





Conclusion Summary

- In the majority of the metro-area counties analyzed, home values increased at a greater rate in 2020 than the average rate from 2017-2019. It also appears that values increased at a proportionally higher level in areas with lower COVID-19 cases.

Implications of findings

What does our conclusion mean?

- According to these metrics, in most areas, we found increased home values in counties surrounding cities with highest COVID-19 cases. This could be due to a desire to leave more densely populated areas, away from areas with higher COVID cases.

What are next steps/further analysis?

- Analyze data from counties with major cities with fewer COVID-19 cases
- Analyze housing values against other market factors to further understand the impact of COVID-19
- Analyze type and size of home purchased against the type of home sold by a particular buyer

Q&A